IBM Data Science Capstone Project Week4

Introduction

This capstone project is part of the requirements to successfully completing the IBM Data Science Professional Certificate. In this project, I am going to focus on Toronto, Canada, where I am going to use all what I have learned throughout this course to suggest suitable locations for a Japanese restaurant based on geographical locations only. Assume that the demand for Japanese cuisines is rising in Toronto, and it becomes a battle of location for prospective Japanese restaurants. I am going to leverage on the Foursquare API to get location data in order to suggest possible gold mine locations for a client who wants to open a Japanese restaurant.

Business Problem

The aim of this project is to suggest suitable locations for a Japanese restaurant in Toronto, Canada. A lot of data science tools and techniques are going to be used along with different machine learning techniques such as clustering. All these are put in place in order to answer the business problem:

"Where would a potential Japanese restaurant owner consider setting up shop in Toronto, Canada?"

Target Audience

In this project, the target audience is an interested potential restaurant owner who wants to open a Japanese restaurant in Toronto, Canada. From the assumption made above, this exercise will predict a good location where it would be easiest to establish a good customer base and patronisers for the restaurant. Think of it as an early bird scenario. This will definitely be of interest to the stakeholder.

Data to be used

In this project, we will need the following data:

1. **A list of neighbourhoods and their postal codes in Toronto, Canada**. This dataset contains three columns which include Postal code, Boroughs and Neighbourhoods. The data than can be extracted from this dataset are rows with a defined Borough.

2. **Geographic coordinates of these neighbourhoods (longitudes and latitudes).** This dataset contains a list of the longitudes and latitudes of the respective postal codes in Canada. We can extract the longitudes and latitudes of the postal codes

related to Toronto so we can be able to work with them when using Folium and Foursquare API

3. **Venue data of Japanese related restaurants in Toronto,** so we can determine the regions with fewer number of Japanese restaurants. This data will contain a list of all the places with in a given radius, so we can use this to get our information of the shops and stores around. The major data we are looking for is the venue category of each shop

Acquiring the data

The list of neighbourhoods will be scrapped from the Wikipedia page of Neighbourhoods in Canada.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,

The geographical coordinates of each neighbourhood will be downloaded from the following link,

http://geoawesomeness.com/developers-up-in-arms-over-google-maps-api-insane-price-hike/,

I have also ensured that this data is authentic.

The Foursquare API will be used to get information on the Japanese related restaurants in the state of Toronto.